5-1-2014

Experimental Philosophy: A Bridge to Progress

Andrew DeLena
Lehigh University

Experimental Philosophy: A Bridge to Progress

Senior Thesis

Andrew De Lena

Lehigh University

Abstract: Experimental philosophy is an emerging methodology which attempts to use empirically gathered data to break new ground on longstanding philosophical debates. Experimental philosophers claim that "armchair" intuitions about philosophical concepts need to be empirically verified among the general population, and that where the two differ the latter cannot be disregarded. I intend to argue that while this approach has its merits it has been plagued by some fundamental misconceptions about the nature of intuitions themselves. Once these misconceptions are addressed, however, I believe experimental philosophy can mature into a fruitful new paradigm, and can achieve its goal of breathing new life into modern philosophy.

INTRODUCTION

Surely everyone would agree that the argumentative form, "Surely everyone would agree that...," is one of the most beloved in all of philosophy. Appeals to what feel like basic, unquestionable, even pre-philosophic intuitions are quite powerful. By showing the reader that each of his or her basic premises is obviously, intuitively correct, a philosopher can construct arguments and theories which will stand up well to criticism. We want to embrace theories that *feel* right. But arguments of this form, appealing to everyone's gut, their intuitions, beg a question which few have ever thought to ask: does everyone share the philosopher's intuitions in the first place? Within the last couple of years an upstart field, calling itself *experimental philosophy*, has begun to ask this question. In a radical departure from the methods of traditional philosophy, experimental philosophers use social psychological methods to discover what beliefs people actually hold on certain philosophical issues. They claim that the intuitions and ideas of John and Jane Doe may not always be identical to those of trained (typically Western) philosophers, but more importantly that John and Jane's intuitions are of vital importance to philosophical theories, and need to be a part of the discussion. Consequently, it threatens some established philosophical beliefs which are either explicitly or implicitly based on appeals to people's intuitions, and has sparked a good deal of controversy in the process. It has had a polarizing effect in its interactions with mainstream modern philosophy, and as such the available metaphilosophical analyses of the field tend to be just as black and white; few unbiased critiques exist of experimental philosophy's merits and flaws. It is my goal to provide such a critique, and in the process to argue for two major points which I believe to be of central importance to experimental philosophy's future success. First, I wish to argue that while the concept of studying laypeople's intuitions is sound, and

potentially useful, a lack of attention to the nature of intuitions themselves, and the processes by which we arrive at them, has been holding experimental philosophy back. Second, I will argue that if experimental philosophy can move past this misconception, or even better, expand its reach beyond the bounds of intuitions, a multitude of new directions for exploration will become available.

In order to give the reader a context in which to view these two arguments, I will first provide a brief history of experimental philosophy and a short description of its typical methodology. I will then outline some of the active debates in the field, followed by what an account of how I believe intuitions have been misperceived and misapplied to these debates. Finally, I will discuss the recent appearance of experimental studies which have moved away from using intuitions as their data, and suggest a few possible directions for future research along these lines.

THE HISTORICAL CONTEXT AND ORIGINS OF EXPERIMENTAL PHILOSOPHY

Experimental philosophy arose from a very real issue not just in modern philosophy, but in the entire history of philosophical thought. The issue: philosophers' reliance on their own *intuitions* as evidence for advancing their respective philosophical theories. From Plato's theory of ideas, to Immanuel Kant's incongruous counterparts, to Thomas Nagel's musings on what it might be like to be a bat, intuitions have furnished evidence and played a central role in advancing philosophical theories for thousands of years.

However, within the last century, the landscape has changed in some interesting ways. Since the beginning of the 20th century, analytic philosophy has radically changed the way most modern philosophy is done, especially in the English-speaking world. Rather than going after a grand theory of the universe in one fell swoop, as had been done by almost every major

philosopher from the Greeks to William James, analytic philosophy takes a divide-and-conquer approach. It evaluates individual, often highly specific issues one at a time. In this regard, analytic philosophy has been heavily influenced by the natural sciences, and accordingly relies heavily upon formal logic and "empirical" investigation. I place "empirical" in scare quotes here because despite the influence of science on analytic philosophy, the two have very different understandings of what counts as empirical evidence. This is the first major point of entry for experimental philosophy. Naturally, what is taken as empirical evidence varies from science to science. Psychology and particle physics, for example, use appreciably different kinds of empirical evidence. But no matter what form it may take, all empirical evidence across the natural and social sciences shares a few fundamental characteristics. First and foremost, it is objective[1]---based either on observation, or, more commonly, experimentation in the physical world. Inherent in this is another, crucial feature of scientific empirical evidence: replicability. A great deal of care is taken in designing scientific experiments such that upon reading a paper, anyone with sufficient expertise could re-run the experiment themselves and achieve the same results. It is not difficult to see how this characterization of empirical evidence differs vastly from that of philosophers. Even among analytic philosophers, the exact same thought experiment can yield vastly different "results" depending on who is thinking about it.[2] Although it takes many shapes, some of which I will discuss later, the most common form of experimental philosophy attempts to settle such disputes by investigating everyday people's intuitions. These, in turn, are used to determine the contours of our folk concepts, such as those of "knowledge" and "intentional action." In the process it aims to explain both intra- and inter-personal variations in intuitions, and reveal any previously unrecognized biases that may be occluding our actual concepts. Stated philosophically, it is a matter of sources

[1] Or, at the very least, able to gain the overwhelming consensus of the scientific community. See Kuhn (1962).
[2] For further discussion of some of the points outlined in this paragraph, see Prinz (2008).

and warrant; experimental philosophy explores the sources of our beliefs and, once it has discovered them, uses this newfound knowledge to determine whether or not those beliefs are warranted (Knobe & Nichols, 2008). Experimental philosophers hope that by uncovering the contours and usages of our folk concepts, they can shift the burden of proof to philosophers who make claims which run contrary to these folk conceptions, and possibly even settle some arguments once and for all.

While it is debatable whether or not analytic philosophy has made philosophical intuitions any more scientifically empirical—intuitions in analytic philosophy still lack objectivity and replicability—it has indisputably changed the types of intuitions to which most philosophers appeal. In essence I believe these new intuitions are still the product of subjective, first-person experience. Intuitions in analytic philosophy are merely couched in high-flung verbiage and technical terms, and tend to be more specific and limited in scope. Experimental philosophy is the most recent response to this change, but the seeds of discontent have been present for decades. In the forward to his 1969 book, *Must We Mean What We Say?*, Stanley Cavell remarks:

> Meaning what one says becomes a matter of making one's sense present to oneself...as though the words we use in philosophy...are *away*...Take, for example, the fact that the isolated analytical article is the common form of philosophical expression now, in the English speaking world of philosophy...This is often interpreted as symptomatic of philosophy's withdrawal from its cultural responsibilities...[we are] ignorant of our cultural situation... (pp. xix-xx).

Clearly Cavell was not referring directly to experimental philosophy, which did not arise for another 30 or so years. But his word choice and depictions of modern philosophy throughout the entire book bear a striking, even uncanny, resemblance to how experimental philosophers have described the need for their methods that exists in the sphere of modern philosophical thought. As Knobe and Nichols (2008) have recently stated, "The rise of analytic philosophy led to a diminished interest in questions about, for example, the fundamental sources of religious faith and

a heightened interest in more technical questions..." (p. 7). They claim that experimental philosophy is a direct response to this sort of shift. As a concrete example, modern philosophy of mind is populated with terms such as *functionalism*, *representationalism*, *heterophenomenology*, and *higher-order global states*. It is hard to get much further *away* from everyday language. During the 1950s, Ordinary Language Philosophy placed a heavy emphasis on just such everyday language, albeit in the form of complex logical analyses of it. However, even the importance of ordinary language in philosophy has faded in and out since then, and has been mostly replaced by efforts to make even more esoteric formal logic the core of language. Wherever intuition is called upon, it is almost always of a shallow, technical sort in the sense described by Nichols and Knobe. The trend has persisted for decades, and is now being attacked by experimental philosophy. Knobe, Nichols, and other like-minded thinkers represent a new breed of philosophers, who seek to reverse this trend of intuitional superficiality and return to discussing more basic, more fundamental questions about the sources of our religious, moral, and metaphysical beliefs (Knobe and Nichols, 2008).

Experimental philosophy is in a somewhat unusual position. It continues this analytical legacy through its emphasis on the methods of modern cognitive science, but is also a refutation of some of analytical philosophy's most fundamental methods, namely the inconsistent, technical, and often unsubstantiated nature of the intuitions it relies upon. The primary goal of experimental philosophy is to shed light on what philosophical intuitions people actually happen to have as means of examining the folk concepts that actually exist in the world, and in so doing to steer the course of modern philosophical debates in a less stagnant, more productive direction. I believe that in theory this sort of approach has its merits. In practice, however, experimental has had some growing pains, most of which center around the concept of intuitions. Before moving on I

will briefly review the prototypical experimental philosophy methodology in order to provide those unfamiliar with the field a context in which to consider these intuitional shortcomings.

THE EXPERIMENTAL PHILOSOPHICAL METHOD

It is probably a misnomer to call experimental philosophy a field of study. Early in its history, its close-knit band of practitioners tended to have converging views on many philosophical issues. As it has developed, however, experimental philosophy has grown to include a huge diversity of standpoints in a multitude of important philosophical domains. Metaphysics, epistemology, ethics, consciousness, metaphilosophy---experimental philosophy has generated important findings in all of these areas. Much more than a field of study, it is a methodology, a unique approach which could potentially be used by a wide variety of people from all sorts of philosophical camps. As Joshua Knobe and Shaun Nichols (2008) aptly state, "What we are proposing is just to add another tool to the philosopher's toolbox. That is, we are proposing another method (on top of all of the ones that already exist) for pursuing certain philosophical inquiries" (p. 10).

So what exactly is the method? Although nothing about the concept of experimental philosophy restricts it to invoking just one scientific field, virtually all experimental philosophy studies conducted thus far have been a hybrid of psychology and modern analytic philosophy[3]. As these two fields have grown over the last 125 years, unfortunately they have grown apart. There is surprisingly little communication between them — philosophers tend to be relatively uninformed about developments in psychology, and psychologists tend not to bother themselves with trying to

[3] As Jesse Prinz (2008) rightly notes, areas like metaphysics, for example, might not have much to gain from the application of psychological methods. But metaphysics could benefit tremendously from increased interaction and understanding between physicists and philosophers. This illustrates the larger point, discussed at length in my final section, of the vast amount of room for expansion in experimental philosophy.

decipher modern theories in philosophy. What's more, the two camps have completely different styles of writing, modes of thought, and theoretical lineages. As part of the growing body of interdisciplinary research brought about by the advent of cognitive science, experimental philosophy attempts to bridge this gap---it applies the findings and methods of experimental psychology to some of the seemingly intractable problems of modern philosophy.

By and large, experimental philosophy's preferred method of empirical investigation is the survey. There are a couple of reasons for this. Jesse Prinz (2008) explains that philosophy's primary method for eliciting intuitions is the thought experiment, and these are relatively easy to convert into surveys. The researcher merely has to simplify the language and ideas involved in the thought experiment so that someone without a philosophy PhD can understand it, ask people what they think about it, and examine the results to divine truths about their concepts and beliefs. In this way, surveys are a natural extension of the traditional philosophical method. They are also relatively simple to design, far less expensive than other modes of psychological research, and participants are much easier to find (typically undergraduate students in introductory level classes). This makes them an additionally logical first step for philosophers without extensive training in experimental design and limited financial and laboratory resources. Studies involving reaction times, behavioral manipulations, and neuroimaging, for example, have not been extensively used by experimental philosophers for pragmatic reasons of this nature, not because they are inherently useless to philosophy. On the contrary, recent experiments which have begun to use such methods are yielding some fascinating results. But setting that aside for later, experimental philosophers, armed merely with their surveys, have been able to gather some very interesting and potentially profound results (Prinz, 2008).

Take, for example, Woolfolk, Doris, and Darley (2006), who conducted a study to investigate how people normally attribute moral responsibility to an agent. Specifically, they hypothesized that two previously overlooked factors play a role in whether or not people will hold a transgressor morally responsible for an action: the extent to which he identifies with his action, and the amount of situational constraint placed upon him. Surprisingly, they found that even when the agent was completely constrained and had no control over his actions, surveys showed that, on average, people still judged him to be morally responsible if he identified with the action he was forced to perform. This presents a serious new issue to be examined in the philosophical debate over the relationship between free will and moral responsibility.

Another good example of a study in a different philosophical arena, namely epistemology, is Weinberg, Nichols, and Stich (2001). Epistemology, broadly defined, is the study of knowledge ---what forms it takes, in what situations it can be acquired, and how we can go about acquiring it. Weinberg, Nichols, and Stich (2001) present evidence which seriously challenges the search for universally normative solutions epistemological problems. They presented classical epistemological thought experiments, in the form of short vignettes, to either Western or East Asian participants and recorded their respective intuitions on those vignettes. For example, in one condition the experimenters used a version of a Gettier case,[4] which read as follows:

> Bob has a friend, Jill, who has driven a Buick for many years. Bob therefore thinks that Jill drives an American car. He is not aware, however, that her Buick has recently been stolen, and he is also not aware that Jill has replaced it with a Pontiac, which is a different kind of American car. Does Bob really know that Jill drives an American car, or does he only believe it?

[4] For a long time epistemologists believed that knowledge was simply justified true belief. However, in 1963 Edmund Gettier challenged that notion by introducing a class of thought experiments where a person has a justified true belief, but we still *feel* like he or she does not have real knowledge. This inconsistency arises because although the subject of the Gettier case has a justified true belief, the way in which he or she acquired it was unanticipated or accidental.

Over 70 percent of Western participants gave what has come to be the standard answer among contemporary philosophers, that Bob "only believes" Jill drives an American car. However, among East Asian participants exactly the opposite pattern was found: the majority believed that Bob "really knows" Jill drives an American car. The experimenters also found major discrepancies in the answers given by participants of different socioeconomic classes, even within the same culture (the same city, actually). This sort of cultural variability in laypeople's pre-theoretical intuitions is a major hurdle for normative epistemological theories---one that may even be insurmountable. If people have different intuitions, then whose intuitions count? That's a tough question to answer.

In addition to their interesting results, I invoke Woolfolk, Doris, and Darley (2006) and Weinberg, Nichols, and Stich (2001) because they are both good illustrations of the way in which experimental philosophy is done, but are very different in from each other in form and content. In the former study, all three authors have a background in experimental psychology[5], and therefore their paper reads very much like something one would find in a peer-reviewed psychology journal. It has a clearly defined introduction and literature review, methods sections, results sections, and discussion. Also, the main issue at hand in the paper---the psychology of moral responsibility---is not in itself all that new for the field of psychology. Many papers have been published on the subject, dating back to the 1970s and beyond (see Hamilton & Sanders, 1981; Darley & Latané, 1969; Lerner & Miller, 1978; for representative examples). The study by Weinberg, Nichols, and Stich, however, differs substantially. It does not read like a psychology journal article, but much more like an analytical philosophy paper. It is not organized by introduction, literature review, etc., but by background, theses, proofs, and objections/replies. In

[5] John Darley, in particular, is actually an exceedingly influential social psychologist. Along with Bibb Latané, he was the first to demonstrate the now-famous "Bystander Effect" in 1969.

addition to the format, Weinberg, Nichols, and Stich explore a subject matter which, to the best of my knowledge, has yet to be explored in the field of psychology. Research on theory of mind comes close, but specific epistemological claims have not previously been tested using experimental methods (Prinz, 2008).

By taking these two papers in conjunction, comparing them side-by-side, one can get a sense for the wide variety of subject matters and argumentative styles which are grouped under the umbrella of experimental philosophy. At one pole there is psychology, at the other contemporary philosophy, and experimental philosophy papers blend the two in varying proportions. But regardless of how they may differ in style and subject matter, the vast majority of experimental philosophy studies conducted thus far share one common goal: to gain insight into the pre-theoretical philosophical intuitions, and ultimately the concepts, of everyday people. This raises the question: why would they be interested in such a thing? A key feature of philosophers' claims to intuitions as evidence for or against a given argument is that their intuitions are shared by essentially everyone (Alexander, 2012). However, no attempt has ever been made by any of these philosophers to empirically verify that this is indeed the case. This is the first goal of experimental philosophy: to verify whether or not the intuitions put forth by various philosophers are actually shared by people everywhere.

A common question that has sprung to many critics' minds goes something like this: "Once these experimentalists discover whether or not a certain philosopher's intuition is shared by the general population or not – what does it matter? What purpose is that newfound knowledge supposed to serve?" This is where various camps in experimental philosophy begin to differ from one another. Alexander, Mallon, and Weinberg (2014) define two major, conflicting views within experimental philosophy on how the intuitions of the general population are to be understood, the

so-called "positive program" and "negative program." The positive program claims that philosophical intuitions are a valuable source of evidence for philosophical theories, and that experimental philosophy is, or should be, and integral part of the process by which we ascertain these intuitions. The negative program, on the other hand, sees experimental philosophy's role as challenging the use, by anyone, of intuitions to advance philosophical theories (Alexander, Mallon, & Weinberg, 2014). Both are relatively radical views. My position is mostly in line with the negative program of experimental philosophy. I do not believe that intuitions, as they are currently conceived, should be used as definitive evidence for any theory. But unsurprisingly, reality is hardly so black and white.

THE AMORPHOUSNESS OF INTUITIONS

Before diving into the gritty details, it may be helpful to briefly outline some of the major debates in experimental philosophy which I will explore, as they pertain to the question of intuitions. In this paper, I invoke a few important lines of inquiry, the first of which is the dispute over the nature of intentional action. As Joshua Knobe (2008) states, "People normally distinguish between behaviors that are performed intentionally, (e.g. raising a glass of wine to one's lips) and those that are performed unintentionally (e.g. spilling the wine all over one's shirt)" (pp. 129-30). But how can we reliably distinguish which actions are intentional and which are unintentional? This is the point of discussion in philosophy. Some, Joshua Knobe for example, believe that moral considerations can actually impact people's judgments on whether or not an action was performed intentionally, even though normatively they probably should not.

If he is correct, this effect is closely tied to the next, and perhaps largest, debate in experimental philosophy thus far: the compatibility of moral responsibility and causal determinism. This has

become an incredibly complex and multifaceted debate, however by and large the discussion in the experimental philosophy community is between so-called Natural Compatibilism (NC) and Natural Incompatibilism (NI). Each of these terms has two parts. Compatibilism, by itself, is the belief that determinism[6] is consistent with both free will and moral responsibility, and that all three can exist in harmony. Incompatibilism, on the other hand, is the belief that determinism is inconsistent with both free will and moral responsibility, and that they cannot co-exist. Natural Compatibilism, then, is the idea that philosophically untrained, normal people have a compatibilist view of the world, and Natural Incompatibilism proposes that people naturally hold the opposing, incompatibilist view.

I will restrict my discussion to the role that intuitions play in studies pertaining to these debates. The first issue I address is a descriptive one. My question is: What *exactly* is forming the data for these experimental philosophy studies? Intuitions, yes—but what are these nebulous and seldom defined feelings, judgments, and dispositions what we lump together and call "intuitions?" The second issue is more prescriptive in nature. In short, given the tremendous amount of variability and outside influence which can affect the formation of intuitions, should we really refer to "intuition" as a single, unified concept at all?

What is an Intuition?

Prior to beginning any scientific experiment, researchers must clearly define the terms used in the experiment. A psychologist, for example, would not begin an experiment which measured participants' reaction times without a clear understanding of what a reaction time is, and how to measure it. In almost all experimental philosophy studies, intuitions are what the experimenter

[6] Alfred Mele defines determinism as the idea that "at any instant exactly one future is compatible with the state of the universe at that instant and the laws of nature" (Mele, 2006 , p. 3).

aims to measure. But what are intuitions? And for that matter, how are they measured? In his book *Experimental Philosophy: An Introduction*, Joshua Alexander covers a number of the competing accounts of intuitions. Some believe philosophical intuitions are "simply beliefs, or perhaps inclinations to believe." Others think they are different from everyday beliefs because they feel "subjectively compelling or necessarily true." A third account claims that philosophical intuitions are "mental states ratified by a process of *philosophical reflection*" (Alexander, 2012, pp. 20-25). These few examples are by no means exhaustive — the list goes on. The important point is the mere fact that there are varying interpretations of what counts as a philosophical intuition. This is clearly an issue for philosophy as a whole, which commonly makes use of intuitions, but it is an especially prominent issue for experimental philosophy since its *raison d'être* is explicitly examining the use and implications of intuitions.

I believe this variability in defining intuitions is a major problem for a few reasons. The self-professed goal of experimental philosophy is to utilize the empirical, scientific methods of experimental psychology. However, if there is no clear-cut definition of what intuitions are, then scientifically measuring them is impossible. The natural first move here would be to attempt to formulate a standard definition of intuitions, perhaps the one that is used most commonly in various experiments. As I will attempt to show, this may or may not be possible. It may be the case that what is universally referred to as our "intuition" is actually a complex and variable blend of other, usually unconscious psychological processes, and that referring to them as one concept does not make sense. In any case, there is a great deal about intuition which we do not yet understand.

Many experimental philosophy papers have touched upon roughly this point, that intuitions are subject to external influence from competing cognitive systems, but to the best of my

knowledge, none have attempted to connect the dots and explain this "influence" systematically. The reasons for this are manifold, but two points by Jesse Prinz (2008) tell most of the story. The vast majority of researchers active in experimental philosophy have a background, and training, in philosophy. Few come from psychology departments, and this leads to a relative lack of interest in the cognitive processes which generate our intuitions. The lion's share of the effort is devoted to accurately describing what the specific lay concepts and intuitions actually *are*. This is the first point. The second deals with the use of surveys. For the reasons previously stated, this has been the primary method of investigation for experimental philosophers so far. But while surveys are useful for collecting intuitions, they are ill-equipped to tackle questions of psychological process which take place below the level of conscious awareness, which I believe is crucial for understanding the nature of the intuitions being collected. Surveys are also ill-equipped to directly folk concepts, which is most experimental philosophers' primary goal. Prinz submits that there are no ideological reasons which make philosophers opposed to directly examining concepts by conducting other forms of experiments, or to examining the cognitive basis of intuitions. It is merely that their philosophical training has made these possibilities less accessible and less salient (Prinz, 2008).

But unfortunately for philosophers who wish to examine folk concepts, it seems that using the current questionnaire methodology there is no way to get around the problem of intuitions, since they are what are being directly measured. Again, this is mostly a practical issue. As Alexander, Mallon, and Weinberg (2014) state, "...this observed variation in intuition would no longer pose a problem if we possessed a means for discerning epistemic wheat from chaff" (p. 38). They argue convincingly that this is impossible through survey research only, and that additional methods are needed in order to formulate such a theory of intuitions. Nevertheless, I do not

believe that survey research has gotten us nowhere. Comprehensive explanatory theories of intuitional variation may be off the table using surveys, but the patterns of responses in various survey-based experimental philosophy studies suggest some very interesting parallels to the reasoning, judgment and decision making, and expertise literatures in psychology, to name a few examples. Let us take a closer look at one such underappreciated parallel between the experimental philosophies of morality and intentionality and the psychology of reasoning.

Possible External Influences on Intuitions

Affective Influence

In perhaps the single most famous experimental philosophy study, Joshua Knobe (2003) examined how people judge that a specific action was intentional versus unintentional. He presented random passers-by in Central Park with one of two scenarios. The first involved an action which brought about a negative side effect:

> The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.' The chairman of the board answered, 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was harmed (p. 191).

The second scenario was logically identical, but involved a positive side effect rather than a negative one:

> The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, and it will also help the environment.' The chairman of the board answered, 'I don't care at all about helping the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was helped (p. 191).

In the negative side effect scenario, 82 percent of participants believed the chairman harmed the environment intentionally, whereas in the positive side effect scenario, only 23 percent of

participants believed he helped the environment intentionally. The two responses are clearly in conflict. Knobe interprets these results to mean that the moral valence of an action plays a role in people's attributions of intentionality; bad actions are more likely to be judged intentional than good ones, and thus application of the lay concept of intentionality is subject to an affective bias. This interference of negative side effects with judgments of intentionality has come to be known as the side-effect effect (Knobe, 2003).

Alexander, Mallon, and Weinberg (2014) argue that this type of inference, that an affective bias is present, requires the ability to individuate the relevant systems which are involved, and that simple survey results do not make possible. However, what if we did not have to derive the contours of the relative systems from Knobe's survey results alone? What if we already had detailed descriptions of such systems, and the question is merely whether or not these results can be explained as instantiations of them in a previously unstudied domain? I believe such a system may be available in the dual-process theory of reasoning. Experimental philosophy papers have posited some sort of dual-process, intuitional and rational, explanation for the results of Knobe (2003) (Cushman & Mele, 2008; Nichols & Knobe, 2008), but to the best of my knowledge only one (Pinillos *et al.*, 2011) has attempted to tie it directly to the dual-process theory in the reasoning literature of psychology. As the name implies, this theory posits two separate systems which people rely upon in different types of scenarios. To use the terms introduced by Stanovich and West (2000), "System 1" is a quick, automatic, heuristic, and often emotional system for reasoning and decision-making, and takes place largely below the level of conscious awareness. Only once the final product of the process has emerged do we become aware of it. "System 2," on the other

hand, is conscious, slow, methodical, and more capable of handling abstract reasoning.[7] It is what philosophers normally think of as the special kind of "human reason" that sets us apart from the rest of the animal kingdom. And, importantly, it is widely thought that System 2 has the ability to override System 1 when the need arises (Evans, 2003). I believe this dual-process model, with its System 1 and System 2, provides a plausible explanation of Knobe's results. Those who used a more System 1 based method of arriving at their intuitions would likely respond the way the majority of Knobe's participants did. When reasoning takes place below the level of conscious awareness, the door is opened for emotion and/or other unidentified psychological influence to bias the reasoning process. It is of no concern to System 1 that the answers which this produces are logically inconsistent. However, those respondents who took the time to stop and use a System 2 sort of method—who overrode their "gut reaction"—would see the inconsistency, and would thus respond in a more instinctively unnatural, but more logically consistent manner. This may be what the 18 percent and 23 percent minorities in Knobe's negative and positive conditions did.

This is all very speculative in nature. Evans (2003) offers a number of demonstrable effects which are typically interpreted as evidence of a dual-process influence. For example, the ability of subjects to see past various biases and provide the logically correct answer to a syllogism has been shown to be correlated with general cognitive ability. Therefore, if it can be shown that higher cognitive ability diminishes the side-effect effect, this would constitute fairly strong evidence that a dual-process effect is at work. There is also evidence that a phenomenon called *disfluency*, where a person gets the feeling that something is amiss with his or her metacognitive processes, can also increase a person's use of System 2 processes (Alter *et al.*, 2007). Again, if this can be shown to

[7] A System 1 method of arriving at intuitions would be similar to Joshua Alexander's (2012) characterization of the Doxastic or Phenomenological Conception of intuitions. A System 2 method would resemble the Etiological or Methodological Conception. So this distinction is recognized by most philosophers, but it is still an open debate as to which is the *best* method. This, as I will discuss, I believe is part of the bigger problem of intuitions.

apply to the side-effect effect, it would strongly indicate the involvement of two separate cognitive reasoning processes. A recent study by Pinillos *et al.* (2011) has found just such results. In separate experiments, they demonstrated that higher levels of intelligence[8] and heightened levels of awareness[9] both significantly reduced the influence of the side-effect effect. The authors believe these results are best understood by appealing to the dual-process theory of reasoning.

Furthermore, it seems the dual-process effect on intuitions is not limited merely to discussions of the side-effect effect. While such direct empirical testing of this claim has not yet been conducted, I believe there is significant anecdotal evidence that warrants further investigation. As an example of this, take Nichols and Knobe (2008), another hugely influential study in experimental philosophy which investigates the Natural Compatibilism (NC) versus Natural Incompatibilism (NI) debate. To recap, NC posits the default, pre-philosophical, layperson's view is that determinism, free will, and moral responsibility are consistent with one another. NI, on the other hand, posits that laypeople naturally believe these concepts are incompatible. Nichols and Knobe attempt to sort out which of these theories, NC or NI, is the correct characterization of the folk concept. In this study, all participants were presented with descriptions of two universes. Universe A was deterministic, and Universe B was indeterministic. The distinction between the two was made clear; participants were told:

> The key difference, then, is that in Universe A every decision is completely caused by what happened before the decision—given the past, each decision *has to happen* the way that it does. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision *does* ***not*** *have to happen* the way that it does. (Nichols & Knobe, 2007, p. 111).

[8] As measured by the number of correct answers (0, 1, 2, or 3) on a Cognitive Reflection Task (CRT) taken after completing a version of the Knobe (2003) paradigmatic CEO experiment (Frederick, 2005).

[9] Pinillos and colleagues used the following logic: "…each CRT question is designed in such a way that one's first pass judgment is mistaken… It is plausible that at this point, you are now made aware that your first pass judgment to problems may very well be mistaken. Accordingly, if you are then immediately given another question, this awareness may then play a role in how you answer that question" (Pinillos *et al.*, 2011).

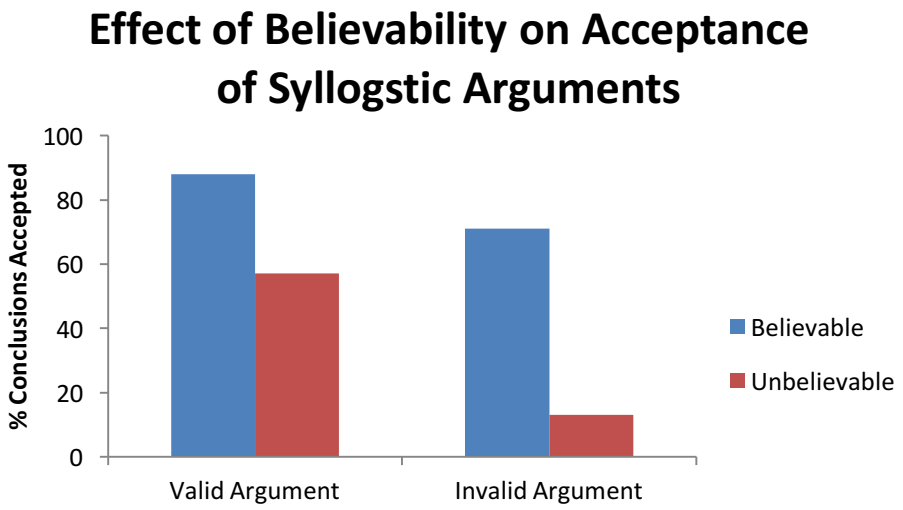
Then, in a 2x2 factorial design (deterministic/indeterministic x high/low affect), participants were randomly assigned to one of the following two conditions:

> **High Affect:** As he has done many times in the past, Bill stalks and rapes a stranger. Is it possible that Bill is fully morally responsible for raping the stranger?
> **Low Affect:** As he has done many times in the past, Mark arranges to cheat on his taxes. Is it possible that Mark is fully morally responsible for cheating on his taxes? (Nichols and Knobe, 2008, p. 117).

Participants were then asked to judge whether Bill and Mark were morally responsible for their actions. As expected, the experimenters found that affect played a significant role in participants' judgments of moral responsibility. In the high affect (Bill) case, participants who were told that Bill was acting in an indeterministic universe judged him to be responsible 95 percent of the time. Interestingly, however, even when told Bill's universe was deterministic, and that his actions *had to happen the way they did*, 64 percent of participants still judged him to be responsible for his actions. So when emotion is high, a majority of participants held Bill responsible for his actions regardless of whether he had free will or not. The pattern of responses in the low affect (Mark) cases was quite different. In an indeterministic universe, 89 percent of participants felt Mark was responsible for cheating on his taxes. But in the deterministic, no free will condition, only 23 percent of people felt he was responsible. Nichols and Knobe interpret this disparity in responses as evidence that the affective content of the scenario greatly impacts people's judgments of moral responsibility. They believe this to be an affective performance error just like the one demonstrated in Knobe (2003), causing people to make non-normative judgments in emotionally charged cases. In emotionally neutral situations, they believe people correctly and competently use their concept of morality to arrive at incompatibilist judgments, making Natural Incompatibilism the correct theory (Nichols & Knobe, 2008).

Although it would take further experimentation to confirm, I believe what Nichols and Knobe term "affective performance error" can, in this case, potentially be viewed as a slight variation on Evans *et al.*'s (1983) belief-bias effect. Evans states, "One of the key methods for demonstrating dual-processes in reasoning tasks involves the so-called 'belief-bias' effect. The methodology introduced by Evans *et al.*...seeks to create a conflict between responses based upon a process of logical reasoning and those derived from prior belief about the truth of conclusions" (Evans, 2003). In studies on this effect, 4 types of syllogisms are typically presented to participants: 1) valid argument, believable conclusion (no conflict), 2) valid argument, unbelievable conclusion (conflicting), 3) invalid argument, believable conclusion (conflicting), and 4) invalid argument, unbelievable conclusion (no conflict). In these experiments, participants are instructed to ignore the content of the syllogism and treat it solely as a logical reasoning task. As it turns out, this is very difficult for participants to do, and the believability of the syllogistic conclusion reliably influences whether or not the syllogism is accepted as valid. Figure 1 displays the typical pattern of responses. Dual-process theory explains these results by proposing that in cases such as these, "although participants attempt to reason logically in accord with the instructions, the influence of prior beliefs is extremely difficult to suppress and effectively competes for control of the responses made" (Evans, 2003).

Figure 1



Source: Evans *et al.* (1983)

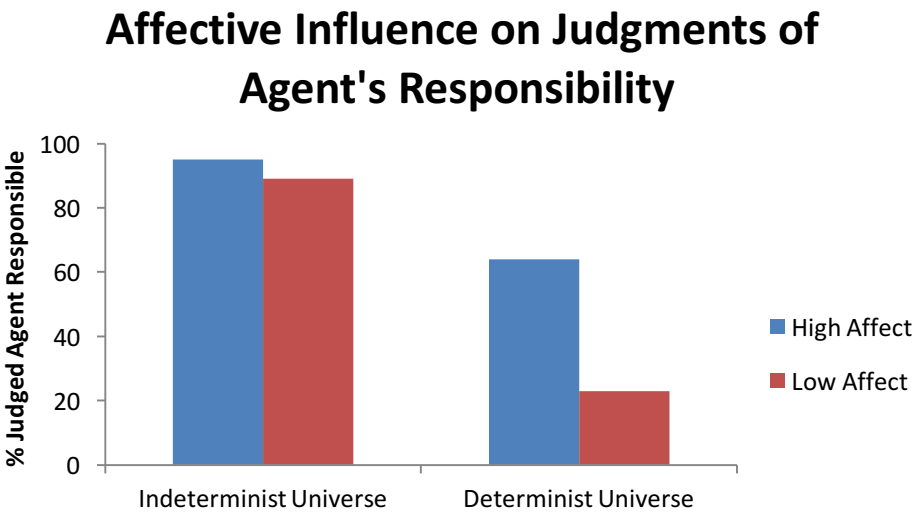
I believe participant in Nichols and Knobe (2008) were operating under the influence of a highly similar effect. It is possible to re-format the experimenters' vignettes into syllogisms similar to those found in Evans *et al.* (1983):

1) Valid argument, believable conclusion (no conflict)
   Bill lives in an indeterminist universe, where he has control of his actions.
   Bill stalks and rapes a stranger.
   Bill is fully morally responsible for stalking and raping a stranger.
2) Valid argument, unbelievable conclusion (conflict)*
   Mark lives in an indeterminist universe, where he has control of his actions.
   Mark cheats on his taxes.
   Mark is fully morally responsible for cheating on his taxes.
3) Invalid argument, believable conclusion (conflict)
   Bill lives in a determinist universe, where he has no control over his actions.
   Bill stalks and rapes a stranger.
   Bill is fully morally responsible for stalking and raping a stranger.
4) Invalid argument, unbelievable conclusion (no conflict)
   Mark lives in a determinist universe, where he has no control over his actions.
   Mark cheats on his taxes.
   Mark is fully morally responsible for cheating on his taxes.

In the case of Nichols and Knobe (2008) it most likely is not the believable/unbelievable distinction that is doing most of the work, but rather, exactly as they suggest, the high/low affect

distinction. Rather than variance in believability of the conclusions themselves, in Nichols and Knobe's study it may have been the strength of participants' desire to believe the conclusions which varied. However, I do not believe this is a crucial point. Emotional thinking plays just as much into System 1 as believability. As a result, when the results of Nichols and Knobe's results are graphed (Figure 2), they look strikingly similar to those of subjects operating under a belief-bias effect. The only place the two patterns differ slightly is in the "valid argument, unbelievable conclusion" condition. This is likely due in part to the fact that the moral nature of these "syllogisms" is more familiar to most people. Consequently, this makes them easier to reason through, not unlike contextualizing the Wason card selection task. Cosmides and Tooby (1992) showed that when this difficult reasoning task, which normally involves following formal rules about letters and numbers, is couched in familiar social contexts, people's performance jumps dramatically. A similar effect could be at play in Nichols and Knobe's "valid, unbelievable" condition. Also, the statement "Mark is fully morally responsible for cheating on his taxes," is not really that unbelievable, nor is it emotionally charged and therefore distorting. It is merely neutral rather than negative. In short, enough features of this condition align with the heuristics of System 1 for people to produce a logically correct judgment most of the time. The other three conditions resemble the classic belief-bias effect much more closely. I believe this is strongly suggestive of a dual-process mechanism—a System 1/System 2 effect—in people's moral reasoning process.

Figure 2



Source: Nichols & Knobe (2008)

Order Effects

In addition to the affective biases demonstrated by Knobe (2003) and Nichols and Knobe (2008), lately another type of intuitional instability has received much attention in the experimental philosophy literature. Order effects, or variation in participant responses depending on the order in which the conditions are presented, have recently been shown to influence intuitions in certain cases, as well. Swain, Alexander, and Weinberg (2008) conducted a study in which they examined people's intuitions on Keith Lehrer's famous *Truetemp* thought experiment, which is typically cited as evidence against reliabilism[10] as a definition for knowledge. The standard thought experiment goes as follows:

> Suppose a person, whom we shall name Mr. Truetemp, undergoes brain surgery by an experimental surgeon who invents a small device which is both a very accurate thermometer and a computational device capable of generating thoughts. The device, call it a tempucomp, is implanted in Truetemp's head so that the very tip of the device, no larger than the head of a pin, sits unnoticed on his scalp and acts as a sensor to transmit

[10] Reliabilism is the idea that a belief counts as knowledge if it is formed or sustained by a reliable cognitive process.

> information about the temperature to the computational system of his brain. This device, in turn, sends a message to his brain causing him to think of the temperature recorded by the external sensor. Assume that the tempucomp is very reliable, and so his thoughts are correct temperature thoughts. All told, this is a reliable belief-forming process. Now imagine, finally, that he has no idea that the tempucomp has been inserted in his brain, is only slightly puzzled about why he thinks so obsessively about the temperature, but never checks a thermometer to determine whether these thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp. Thus, he thinks and accepts that the temperature is 104 degrees. It is. Does he know that it is? (Lehrer, 1990, pp. 163-164)

According to reliabilism, Mr. Truetemp does know that it is 104 degrees since his belief is caused by a reliable cognitive process. But most people have the intuition that Mr. Truetemp, for some reason, does *not* really know that it is 104 degrees, and this is supposed to count as evidence against reliabilism.

Swain and colleagues took a different approach to this thought experiment. In their study, they utilized four different thought experiments, varying only the order in which they were presented. One thought experiment presented a clear case of knowledge, involving a chemist who correctly realized that the mixing of two chemicals produces a toxic gas. Another presented a clear case of non-knowledge, in which a man sometimes gets a "special feeling" before flipping a coin, and when he gets this special feeling he can predict the outcome of the coin flip about half of the time. The third thought experiment was an uncertain case of knowledge, where a woman unknowingly driving through a countryside movie set sees the only real barn among a whole population of fake barns. The fourth was the *Truetemp* case. Swain, Alexander, and Weinberg hypothesized that when first presented with a clear case of knowledge (i.e. the chemist case), participants would be less likely to attribute knowledge to Mr. Truetemp. Conversely, when first presented with a clear case of non-knowledge (correctly predicting 50 percent of coin flips), they would be more likely to attribute knowledge to Mr. Truetemp. This is exactly what they found. The authors take this to be especially important because in Lehrer's book, the chapter in which

the *Truetemp* case is introduced immediately follows a chapter on paradigm cases of knowledge, such as perceptual knowledge and knowledge of mathematics. They believe this has a similar effect as their *Chemist* case, making the reader less likely to attribute knowledge to Mr. Truetemp. This is an especially clear example of how a lack of understanding of intuitions can have profound consequences—many factors other than the folk concept of knowledge are clearly at play here. A final point was also interesting: people's intuitions on the fake barn case, which was also an uncertain case of knowledge, remained surprisingly stable regardless of the order of presentation. This raises the additional question of how to know *a priori* which intuitions are susceptible to biasing effects and which are not (Swain, Alexander, & Weinberg, 2008).

Similarly to affective biases, order effects have been shown to impact intuition formation in multiple different areas of philosophical discussion. Cushman and Mele (2008), for example, return to the question of intentional action. The authors borrowed the thought experiment about the CEO harming or helping the environment from Knobe (2003), and also generated 15 more of their own, similar cases. Approximately half of the subjects saw the CEO harm case within the first four cases presented; the other half saw it within the last four cases presented. They found that, "subjects were five times as likely to judge that the CEO did not intentionally harm the environment when responding toward the end of the test, compared to subjects who responded to CEO harm towards the beginning" (Cushman & Mele, 2008, p. 176). Whatever the reason for this, it is clear that something outside the thought experiment itself is influencing people's judgments about it. This is a major problem for experimental philosophers. Any inference from survey results to the actual folk concepts which are supposedly behind them runs the risk of overlooking order effects, affective biases, or any number of as yet undiscovered influences which may be getting in the way.

I have reviewed only a handful of the various studies experimental philosophy studies which point to seemingly non-normative, outside influences on people's intuitions. In addition to the intrapersonal instability described in the preceding section, there is a wealth of equally fascinating studies on patterns of interpersonal differences. Responses vary between cultures (Machery, Mallon, Nichols, & Stich, 2008; Weinberg, Nichols, & Stich, 2001), between gender (Zamzow & Nichols, 2009; Buckwalter & Stich, 2014), even between socioeconomic classes (Weinberg, Nichols, & Stich, 2001), none of which should normatively have any bearing on universal questions of truth, right and wrong, or the nature of consciousness, to give a few examples.

Some Hope for Intuitions?

So, what does all this instability and variability mean for experimental philosophers? Are intuitions totally epistemically useless, or is there some way they can be saved? This is a topic of spirited debate in the experimental philosophy literature. On the one side, there those like Stacey Swain and Jonathan Weinberg who believe intuitions are far too open to outside, normatively irrelevant factors, and philosophy should do away with appealing to them altogether. On the other side, there are some equally interesting defenses of intuitions. Recall that the most worrying point made by Swain, Alexander, and Weinberg is not that a wide swath of our intuitions may potentially be unstable, but that we have no way of predicting which intuitions these will be. Wright (2014) argues that this is not true—we do, in fact, have ways of predicting the stability of our intuitions. In two studies, Wright showed that confidence, belief strength, and perceived paradigmaticity were all significant predictors of the stability of a particular intuition. People's confidence in cases like *Truetemp* tended to be low, and thus subject to the order effects demonstrated by Swain and

colleagues. But confidence and strength of belief tended to be quite high in cases like the *Chemist*, and these cases showed no order effects at all. If Wright is correct, and there is a way to *a priori* predict the extent to which an intuition is subject to bias, this is certainly good news for traditional philosophers and positive-program experimental philosophers.

In the end, however, I doubt the usefulness of confidence and belief strength in the problem of intuitions. They may be able to predict which intuitions will be stable and which will be unstable; however, in general, those intuitions which are the most stable tend to be the least philosophically interesting. I do not believe that clear-cut cases of knowledge, like the *Chemist*, provide us with much information about the concept of knowledge itself. If we want to know the size and shape of the concept of knowledge, we have to find its boundaries. This means venturing into and embracing the unstable territory of cases like *Truetemp*, with all its order effects, affective biases, etc. However, because these questions are so complex and unclear, in order to explore them experimental philosophers need more sophisticated tools than surveys to properly do the job. We must be able to pull apart the various systems and other factors which play a role in the formation of our intuitions, and surveys are ill-equipped for this task. In short, if experimental philosophers wish to find the borders of concepts such as knowledge, I believe they must explore the fringes of these concepts with more resolution and specificity than intuition-based surveys can provide.

My View

In many ways, the debate over the epistemic value of intuitions is the debate over the future of experimental philosophy, or at least what form it will take. But regardless of which side of the fence one comes down upon—the positive program or the negative program—it is fairly clear that

something has to change. Minor qualms with certain methodologies aside, the experimental evidence is pretty undeniable: the traditional method of appealing to intuitions as evidentiary support for philosophical theories seems to be untenable. So, what is the correct way to view intuitions? And given this correct view, how must the traditional philosophical method change to accommodate it?

I believe there are two strong candidates for the correct way to view intuitions. First, it could be that experimental philosophy is more or less on the right track; intuitions could be the end product of a variable collection of cognitive processes. As an analogy, consider the everyday act of driving, for instance to cousin Bobby's house. Say my sister and I start from the same house, but we each take a different route to get to Bobby's. So we start in the same place and end in the same place, but take different paths in between. If this is the case with intuition, then it makes sense to continue to refer to intuition as we always have, as one concept. The challenge for experimental philosophy, then, is to start from "Bobby's house" (intuition) and re-trace the routes my sister and I both took to get there, which will lead them to our "home" (underlying concept). This is a very crude analogy, but I hope it illustrates the point.

I believe there is also a second, slightly more radical possibility. It could be that the correct way to view intuitions is to stop viewing them as intuitions—to stop viewing them as one unitary concept that must be defined in one and only one way. To use the same analogy as the previous case, intuitions would be more like "cousins" than "cousin." My sister and I would start from the same house, take different roads, but I end up at cousin Bobby's house and she ends up at cousin Billy's house. Intuitions on moral responsibility, for example, may be completely different in nature than intuitions on consciousness, or intuitions on free will. On this view, not only would we

arrive at our intuitions in various ways, but the end products are inseparable from the paths by which we arrive at them, causing the end products to be different in kind, as well.

Surprisingly to me, it seems that while this second view has been half-acknowledged in some papers, no one has ever seriously considered this possibility. To the best of my knowledge, multiple kinds of intuitions have only been considered as competitors to one another. Weinberg *et al.* (2001) distinguish between different kinds of intuitions in their replies to possible objections, essentially delineated by how much reflective thought they involve. There is clearly the implication, however, that some intuitions are better than others. Also, as previously mentioned, Joshua Alexander (2012) addresses the issue more explicitly, outlining five different competing definitions of intuitions. There is the *Doxastic Conception*, which treats intuition as a synonym of belief, all the way up through the *Methodological Conception*, which treats them as, "mental states ratified by a process of *philosophical reflection*" (p. 25). Each of them have their strengths and weaknesses, and are endorsed by different philosophers. But none of them seem to cover all that we call *intuition* (Alexander, 2012).

Alexander presents these various accounts as competitors. However, if we were to stop viewing intuition as one entity, they may not necessarily have to be. My speculation is that all of the conceptions presented by Alexander can be found in philosophy, maybe even within experimental philosophy. As a brief example, recall the Nichols and Knobe (2008) experiment on Natural Compatibilism (NC) versus Natural Incompatibilism (NI) (described on pp. 18-19 of this paper), which found that emotionally charged scenarios elicit NC intuitions and emotionally neutral scenarios elicit NI intuitions. Feltz, Cokely, and Nadelhoffer (2009) conducted a follow-up study which at first seems to challenge Nichols and Knobe (2008). There was only one difference between the two studies: Nichols and Knobe utilized a between-subjects experimental design,

while Feltz, Cokely, and Nadelhoffer utilized a within-subjects design. In a between-subjects design, participants only see one of the four possible scenarios (deterministic/intedeterministic x high/low affect). In a within-subjects design, however, participants saw all four possible scenarios and could compare them side-by-side. As a result, in Feltz, Cokely, and Nadelhoffer (2009), participants gave much more stable, consistent answers than those in Nichols and Knobe (2008). If we allow for the co-existence of multiple conceptions of intuitions, then the results of these two studies are not in conflict, as the authors believe them to be. Because of the experimental design, Nichols and Knobe's participants may have primarily utilized a doxastic-conception type of intuition, while Feltz, Cokely, and Nadelhoffer's participants may have primarily utilized a methodological-conception type of intuition. Each may be able to tell us something different about the psychology involved in the complex debate between NC and NI.

At this point it is too early to say with confidence which view, the intuition-as-end-product view or the intuition-as-conflation view, is closest to the truth. Regardless, I believe more investigation into intuitions themselves is needed; not just as windows to our folk philosophical concepts, but in their own right. A deeper understanding and more fine-grained vocabulary of cognitive processes that play a role in the formation of philosophical beliefs would breathe some fresh air into modern philosophy, and bring it closer to the empirical (and in some cases purely descriptivist, as I will explain) field of study which analytic philosophers always wanted it to be.

AN EVOLVING METHODOLOGY

If experimental philosophers can force this kind of a change, I think they can make an incredibly valuable contribution to the future of philosophy. That being said, in order to do this they will need to make some significant changes, themselves. Anyone who does work in the field

of experimental philosophy needs to possess not only knowledge of empirical methods and statistics, but also knowledge of the relevant psychological (or other scientific) literature which relates to their area of investigation. I believe the dual-process theory of reasoning, for example, should have been invoked in discussions of affective bias from the beginning. This once again raises the other point made by Jesse Prinz (2008), that the main focus of experimental philosophy has been on laypeople's *concepts*, not the processes by which they generate them. He argues that to explore this latter question, a deeper understanding of psychology and its methods is needed within experimental philosophy. When coupled with arguments such as those in Alexander, Mallon, and Weinberg (2014) and Kauppinen (2007), however, one consistent theme emerges: surveys are not the answer. If experimental philosophy is to make significant contributions to our understanding of philosophical concepts and thought processes, it needs to evolve and start using more genuinely psychological methodologies. Admittedly, this does blur the lines between philosophy and psychology to a considerable degree. But similar to my views on Wright's (2014) experiment, I believe gray area between what we normally consider to be philosophy and psychology is probably more interesting and more informative for the question of intuitions than the traditional core of each discipline. Even if this sort of experimental philosophy which I am proposing does cross the border and become bona-fide cognitive psychology, I still believe it is a necessary step which experimental philosophy must take if it wishes to achieve its original goal of understanding our folk philosophical concepts. The "black box" of survey-gathered intuitions currently prevents this sort of understanding, but if experimental philosophy can utilize other psychological methods, it may really start to get somewhere.

Auspiciously, such changes are already starting to take place. The recently published *Experimental Philosophy: Volume 2* (Nichols & Knobe, 2014), for example, includes a reaction

time study by Arico *et al.* (2014), in which they demonstrate that people are slower to deny consciousness to entities that merely possess simple characteristics like eyes and interactive behavior. This implies that the recognition of an entity as an *agent*, which the above characteristics trigger, plays a significant role in the attribution of conscious states. Young, Nichols, and Saxe (2010) used behavioral and fMRI methods to investigate cases where "bad luck" seems to influence moral judgments. They show that the major factor in cases involving moral luck is not whether an agent's action results in a harmful outcome, but whether or not people feel the agent was justified in believing that his action would not cause a harmful outcome. Studies such as these eliminate much of the ambiguity that plagues survey studies, and, I believe, are blazing a trail which the rest of experimental philosophy should follow.

DIRECTIONS FOR FUTURE EXPLORATION

A shift in methodology—away from surveys and towards other, more commonly used techniques—would not only improve the quality of experimental philosophy results, but would also allow it to explore previously untapped areas of philosophy. The overwhelming majority of experimental philosophy studies conducted thus far have been in the fields of epistemology, action, and ethics, with a few forays into the philosophy of mind and metaphysics. This is all well and good, experimental philosophy can make valuable contributions to these areas if conducted in the right way. But I believe the canon of Western philosophy is replete with questions that lend themselves to empirical investigation, even in some unlikely places. I believe empirical investigations of some the writings of Wittgenstein, Cavell, and Foucault, as just a few examples, could yield some fascinating results.

Wittgenstein

Ludwig Wittgenstein was decidedly anti-empirical. For this reason, I believe his Philosophical Investigations (2009) (PI) presents an interesting challenge to experimental philosophy, mainly for two reasons. First, it first poses a challenge to experimental philosophy's mere existence, and second, if the first challenge can be overcome, I believe it opens up a great deal of Wittgenstein's writings to empirical exploration. Wittgenstein himself was diametrically opposed to this idea of philosophy as science. In the PI he writes, "It was correct to think that our considerations must not be scientific ones...And we may not advance any kind of theory. There must be nothing hypothetical in our considerations. All *explanation* must disappear, and description alone must take its place" (§109). Somewhat unexpectedly, this, by itself, fits well with my own personal vision for experimental philosophy. I believe it is most valuable as a descriptive project, just as Wittgenstein believed for philosophy as a whole, but if it ventures into theory it may find itself on shaky ground. Experimental philosophy's value, to borrow Wittgenstein's words, is that it reminds us that, "The aspects of things that are most important for us are hidden because of their simplicity and familiarity. (One is unable to notice something—because it is always before one's eyes)" (§129). Wittgenstein believed that by paying careful attention to the language we use in everyday situations, we could discover these hidden features of life. In a certain sense I agree completely; there is much to be discovered in the language we take for granted. That being said, in working primarily with European languages and cultures Wittgenstein may have had an incomplete picture. This incompleteness is revealed by none other than experimental philosophy. In support of his argument against the idea of private language, Wittgenstein says the following: "In what sense are my sensations *private*? – Well, only I can know whether I am really in pain; another person can only surmise it. – In one way this is false, and in another nonsense. If we are using the word "know" as it is normally used (and how else are we to use it?), then other people very often know if

I'm in pain" (§246). Empirical epistemology studies, however, have given us fairly convincing evidence that the ordinary usage of the word "know" depends on where you are and who you ask (Weinberg, Nichols, & Stich, 2001; Woolfolk, Doris, & Darley, 2006). This is not to say that Wittgenstein's remarks are necessarily invalidated by such findings. It may be the case that once we have a complete picture of the ordinary usage of the word "know," Wittgenstein's invocation of it will remain perfectly valid. At this point, however, we cannot know for sure. In order to be sure, I think experimental philosophy's priority should be mapping out the usage of "know" and other major concepts like it. Lest anyone think that removing the theorizing element reduces its status, or makes it just another social science, Wittgenstein, once again, puts it in perspective, "...this description gets its light – that is to say, its purpose – from the philosophical problems...The problems are solved, not by coming up with new discoveries, but by assembling what we have long been familiar with. Philosophy is a struggle against the bewitchment of our understanding by the resources our language" (§109).

Foucault

This emphasis on description as the primary purpose of philosophy is made even clearer by the writings of Michel Foucault. Whereas Wittgenstein was concerned primarily with language, Foucault was a philologist and a philosopher of politics, among other things. Describing his own work, Foucault writes in his *Dits et écrits*, "Very schematically, it is this: to try to recover in the history of science, of knowledges [*connaissances*] and of human knowledge [*savoir humain*] something that would be like the unconscious of it...that would have its own rules, as the unconscious of the individual human being also has its rules and its determinations" (Davidson, 1997, p. 7). Already the similarities are striking between Foucault and this new experimental philosophy paradigm. Experimental philosophy has mainly focused on individuals, their concepts,

and the processes by which they form them; it has only invoked culture to the extent that it explains these patterns. However, it is not difficult to see how its research program could be expanded to a macro scale, studying societies and civilizations as a whole. Joshua Knobe practically says as much in his "Experimental Philosophy Manifesto." He claims that for most of the history of philosophy, "it wasn't particularly important to keep philosophy clearly distinct from psychology, history, or political science. Philosophers were concerned, in a very general way, with questions about how everything fit together. The new movement of *experimental philosophy* seeks a return to this traditional vision" (Knobe, 2008). Foucault's writings on language, politics, and society are practically a tailor-made entry point for experimental philosophy. One instance of this is Foucault's inquiry into the relationship between structural linguistics and society. Speaking on these ideas, Arnold Davidson writes,

> Given this type of analysis, the important empirical question arises, 'Up to what point can relations of a linguistic type be applied to other domains and what are these other domains to which they can be transposed?' But Foucault turns directly to a second question... are these kinds of relations, discovered by structural linguistics and perhaps extendable (this is the first empirical question) to myths, narratives, kinship, and society in general, capable of being completely formalized? (pp. 8-9).

I think this would make for a fascinating mix of experimental and historical research, and is perfectly suited for the social psychological methods experimental philosophy employs. It may also give experimental philosophy a chance to defend itself against some of the claims made by Wittgenstein, who believed that linguistic relations held all the answers to the societal and epistemological questions listed by Davidson. Foucault probably also believed this, but he was more skeptical; he did not believe we could assume as much *a priori*. This difference is crucial—it is the difference between experimental philosophy being useful, or useless. The safe bet, in my mind, is to side with Foucault. We may discover Wittgenstein is right, and we were simply overly

cautious. It would be a shame, however, to assume as much from the start and potentially miss something of major importance.

Despite this one major difference between Wittgenstein and Foucault, in many other ways they had remarkably similar views, to the benefit, I think, of experimental philosophy. A specific example of this can be found in Foucault's ideas on relations of power—not only in the sense of international economics or political hegemons, but also the power of paradigms and ideas. If nothing else, a better understanding of this would allow philosophy to view itself in the mirror. Channeling Wittgenstein's idea that "One is unable to notice something—because it is always before one's eyes"(§129), Foucault claimed, "'the task of philosophy today could well be, What are these relations of power in which we are caught and in which philosophy itself, for at least one hundred and fifty years, has been entangled?'" (Davidson, 1997, p. 3). If "analytic philosophy" is substituted for "relations of power," the resemblance to experimental philosophy is striking. Historically, and on a high level, we may have a rough approximation of how this analytical paradigm came to be dominant. The details of this usurpation, however, and the specific ways in which it guides and shapes our thinking, are almost completely unexplored. Experimental philosophy seems to me just the way to investigate these and other questions of power, in all its forms. After all, psychology has already made some shocking discoveries on the power of conformity (Asch, 1951) and authority (Milgram, 1974), for example. But in order to pursue Foucault's questions, experimental philosophy will need to make some revisions to its research paradigm. It cannot approach these questions of power in the same way it has approached intuitions. Instead, as with intuitions, a different focus is needed. Davidson, summarizing a passage by Foucault, says, "...we should not assume that relations of power have only one function; we should describe power, in all of its diversity and specificity, as it actually works" (p. 4).

Experimental philosophy needs to see past the traditional terminology and modes of thought, which are so pervasive they have become invisible, and become the truly descriptive project which Wittgenstein and Foucault sought.

Cavell

Stanley Cavell's writing has such an experimentalist flavor to it that, frankly, I am shocked no one in experimental philosophy has thought to call upon him before. At any rate, the reason I do so here is, fittingly, given by Cavell himself when he says of Wittgenstein, "I find that his *Philosophical Investigations* often fails to make clear the particular way in which his examples and precepts are to lead to particular, concrete exercises and answers" (Cavell, 1969, p. XXV). Cavell provides us with concrete exercises and answers which may help to jump start inquiry in the new direction suggested above. Chapter 1 of *Must We Mean What We Say* deals largely with the above-mentioned issue of whether or not empirical investigation is necessary to affirm philosophical truths. Wittgenstein did not believe it was; I think Foucault did. Cavell falls somewhere in between, but in the process provides specific cases that could be useful in determining who was closest to the truth. In his introduction of a debate between Gilbert Ryle and Benson Mates, which is highly similar to the difference I have framed between Wittgenstein and Foucault, Cavell writes, "One of Mates' objections to Ryle can be put this way: Ryle *is* without evidence—anyway, without very good evidence—because he is not entitled to a statement of the first type (one which presents an *instance* of what we say) in the absence of experimental studies which demonstrate its occurrence in the language" (p. 4). This might at first sound like a resounding (if anachronistic) endorsement of experimental philosophy, but Cavell quickly explains, "To answer *some* kinds of specific questions, we will have to engage in that 'laborious questioning' Mates insists upon, and count noses; but in general, to tell what is and isn't English, and to tell whether what is

said is properly used, the native speaker can rely upon his own nose; if not, there would be nothing to count" (p. 4). This statement is interesting for a number of reasons. On the one hand, it is certainly true in some cases, and therefore lends support to the Wittgensteinian argument. As a competent speaker of English, I am entitled to say "It is currently raining," without taking a survey to verify that my fellow English speakers agree. In fact, doing so might even confuse the issue by giving undue weight to the portion of people who would disagree because they have not been outside, have not seen the forecast, etc.; namely, the people who are subject to various performance errors, as was most likely the case in the thought experiment on the CEO intentionally helping or hurting the environment (Knobe, 2003).

Statements on the weather are a pretty mundane example—Cavell provides more interesting ones. A few pages after its initial introduction, he further explains his belief that competent speakers of a language have a right to claim "what we should mean in (by) saying [words]" using the example of the word "voluntarily" (p. 8). According to Cavell, we only use "voluntarily" when something about a situation or an action is "fishy," like if I was to ask someone, "Did you wear that hat voluntarily?" Here, nothing about the denotative meaning of "voluntary" conveys this fishy sentiment, but it is an undeniable part of the word's usage in ordinary language. Cavell says we understand this use of "voluntary" by virtue of being English speakers, and that no counting of noses is required. This, according to him, is the normal state of affairs in the ordinary use of language. In order for them to be justifiably questioned, there must be something weird going on. He says, "My point about such statements, then, is that they are sensibly questioned only when there is some special reason for supposing what I say about what I...say to be wrong; only here is the request for evidence competent (p. 14).

This sounds about right, as far as it goes. But I think experimental philosophy has given us at least some cause to doubt the universality of Cavell's claims. At one point, he writes, "I am prepared to conclude that the philosopher who proceeds from ordinary language is entitled, without special empirical investigation, to...assertions like, 'We do not say "I know..." unless we mean that we have great confidence...," and like, "When we ask whether an action is voluntary we imply that the action is fishy'" (p. 12). However, I do not believe the first and the second example are equivalent. The case of "voluntary" is much more limited in scope than the case of "know," and I think this introduces complexities which make the two significantly different. For instance, I do not think that if Swain, Alexander, and Weinberg were to re-run their order effects experiment, examining the concept of "voluntary" instead of "know," would they attain the same results. "Know" is a much more versatile term—the pragmatics are much more complex, to use Cavell's words—and this complexity introduces certain scenarios in which the usage of "know" is influenced by things like culture, emotion, order effects, etc. So where do we draw the line on what we have a right to claim *a priori* and what must undergo experimental examination? This is hard question to answer. Ultimately, actually doing experiments, using a trial-and-error approach, may be the only way to find out.

Summary

In this section, I have tried to show, using a few specific examples, that the history of philosophy is riddled with claims and ideas which could benefit from experimental investigation. The three philosophers I have invoked here are not by any means exhaustive, but they provide some interesting beginning points which have not yet been addressed by experimental philosophy. To reiterate, Wittgenstein poses an interesting challenge to experimental philosophy in that he was a staunch opponent of the idea that empirical methods were necessary in philosophy. He believed

that formal analysis of language could reveal rules and relationships that generalize to all aspects of life. But Wittgenstein also saw philosophy's role as being descriptive, not hypothetical or explanatory. Foucault was similar in many ways; he believed that structural linguistics could shed much light on culture and history. He believed, however, that this approach could only go so far. At a certain point we must get our hands dirty and investigate relations of power, for example, in their own right. It takes a certain degree of experimental study to get at unconscious structures of individuals' thought, and the same applies to societies. Foucault thought, and I agree, that there is a great deal of value in understanding the gritty details of these unconscious structures—they allow us to truly understand the forces that shape our thoughts and actions. It is a tall order, but I believe experimental philosophy is in a great position to go after just this kind of global description. Finally, Cavell falls somewhere in between. Some of his examples show that empirical study has its limits, and it not always warranted. Yet sometimes knowingly, sometimes unknowingly, leaves room for empirical study where Wittgenstein did not. I do not claim to have any sort of conception on where to draw the line for experimental philosophy, and by extension empirical science as a whole, but finding that line could potentially be of great value in its own right.

CONCLUSION

I hope I have been able to convey my belief that there is a very real need for experimental philosophy in the greater picture of modern philosophy, and modern psychology, too. To quote Jesse Prinz (2008), echoing Kant: "Data without theory is empty, and theory without data is blind" (p. 205). I believe experimental philosophy can provide a healthy dose of theory to relevant parts of psychology, and data to philosophy; provided it is conducted in a genuinely scientific way. The field is young, and it has made progress, but experimental philosophy needs to continue to transition away from the use of primitive questionnaires in favor of more suitable research

methods. What those methods are in practice will depend in part on the topic at hand. In pursuing questions of consciousness and philosophy of mind, fMRI and other functional imaging techniques may be useful tools. For epistemological and moral studies, reaction time and behavioral experiments may be a good choice. A global theory of Foucault's unconscious structures of power may require all of the above, in combination with historical research. It is difficult to predict exactly what will be needed, but experimental philosophy cannot continue to shy away from the most appropriate methodologies due to a lack of expertise. In the short term, I believe more active collaboration between experimental philosophers and psychology departments will greatly increase the quality of experimental studies. This would almost certainly help to bring a more balanced focus to experimental philosophy, bringing the processes by which we form and use concepts up to an equal standing with the concepts themselves. In the long term, I hope more and more contributors to the field will represent a "new breed" of cognitive scientists, trained in multiple disciplines from the start. Again, I believe experimental philosophy is making progress in this direction, but there is much more progress yet to be made. Personally, I think this is an exciting prospect. So far, experimental philosophy's main quest has been to define the space of intuitions, and the concepts the accompany them. But if experimental philosophy can lead the way in breaking through the traditional mold of intuitions, entire new vistas will appear before philosophy—vistas that have always been invisible because they were right before our eyes.

Works Cited

Alexander, J. (2012). *Experimental Philosophy: An Introduction.* Cambridge: Polity Press.

Alexander, J., Mallon, R., & Weinberg, J. M. (2014). "Accentuate the Negative." In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy: Volume 2* (pp. 31-50). Oxford: Oxford University Press.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007) Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology, 136,* 569–576.

Arico, A., Fiala, B., Goldberg, R. F., & Nichols, S. (2014). "The Folk Psychology of Consciousness." In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy: Volume 2* (pp. 111-136). Oxford: Oxford University Press.

Buckwalter, W., & Stich, S., (2014). "Gender and Philosophical Intuition." In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy: Volume 2* (pp. 307-346). Oxford: Oxford University Press.

Cavell, S. (1969). *Must We Mean What We Say?* Cambridge: Cambridge University Press.

Cosmides, L., & Tooby, J. (1992). "Cognitive Adaptations for Social Exchange." In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The Adapted Mind: Evolutionary psychology and the generation of culture* (pp. 163-228). New York: Oxford University Press.

Cushman, F., & Mele, A. (2008). "Intentional Action: Two-and-a-Half Folk Concepts?" In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy* (pp. 171-188). Oxford: Oxford University Press.

Davidson, A.I. (1997). "Structures and Strategies of Discourse: Remarks Towards a History of Foucault's Philosophy of Language." In A.I. Davidson (ed.), *Foucault and his Interlocutors*, (pp. 1-17). Chicago: The University of Chicago Press.

Evans, J.St.B.T (2003). In two minds: Dual-process accounts of reasoning. *TRENDS in Cognitive Sciences, 7,* 454-459.

Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11,* 295–306.

Feltz, A., Cokely, E. T., & Nadelhoffer, T. (2009). Natural Compatibilism versus Natural Incompatibilism: Back to the Drawing Board. *Mind & Language, 24,* 1-23.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108,* 814-834.

Hamilton, V. L., & Sanders, J. (1981). The Effect of Roles and Deeds on Responsibility Judgments: The Normative Structure of Wrongdoing. *Social Psychology Quarterly, 44,* 237-254.

Kauppinen, A. (2007). The Rise and Fall of Experimental Philosophy. *Philosophical Explorations, 10,* 95-118.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis,63,* 190-194.

Knobe, J. (2008). "The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology." In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy* (pp. 129-148). Oxford: Oxford University Press.

Knobe, J., & Nichols, S. (2008). "An Experimental Philosophy Manifesto." In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy* (pp. 3-14). Oxford: Oxford University Press.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Latané, B., & Darley, J. M. (1969). Bystanders "apathy". *American Scientist,57,* 244-268.

Lehrer, K. (1990). *Theory of Knowledge*. Boulder: Westview Press.

Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin, 85,* 1030-1051.

Machery, E., Mallon, R., Nichols, S., & Stich, S.P. (2008). "Semantics, Cross-Cultural Style." In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy* (pp. 47-60). Oxford: Oxford University Press.

Mele , A. (2006). *Free Will and Luck*. Oxford: Oxford University Press.

Nichols, S., & Knobe, J. (2008). "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions." In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy* (pp. 105-126). Oxford: Oxford University Press.

Pinillos, N. Á., Smith, N., Nair, G. S., Marchetto, P., & Mun, C. (2011). Philosophy's New Challenge: Experiments and Intentional Action. *Mind & Language, 26,* 115-139.

Prinz, J. J. (2008). "Empirical Philosophy and Experimental Philosophy." In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy* (pp. 189-208). Oxford: Oxford University Press.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23,* 645-726.

Swain, S., Alexander, J., & Weinberg, J. M. (2008). The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp. *Philosophy and Phenomenological Research, 76*, 138-155.

Weinberg, J. M., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics, 29,* 429-460.

Wittgenstein, L. (2009). *Philosophical Investigations.* (4th ed.). Chichester: Wiley-Blackwell.

Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition, 100,* 283-301.

Wright, J.C. (2014). "On Intuitional Stability: The Clear, the Strong, and the Paradigmatic." In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy: Volume 2* (pp. 51-74). Oxford: Oxford University Press.

Young, L., Nichols, S., & Saxe, R. (2010). Investigating the Neural and Cognitive Basis of Moral Luck: It's Not What You Do but What You Know. *The Review of Philosophy and Psychology, 1,* 333-349.

Zamzow, J. L., & Nichols, S. (2009). Variations in Ethical Intuitions. *Philosophical Issues, 19,* 368-388.